

友人の影響を考慮した行動予測

矢萩 総太[†] 松林 達史^{††}

[†] 電気通信大学大学院 情報理工学研究所 情報学科 〒182-8585 東京都調布市調布ヶ丘 1-5-1

^{††} 日本電信電話株式会社 NTT サービスエボリューション研究所 〒239-00847 神奈川県横須賀市光の丘 1-1

E-mail: [†]syahagi@ni.is.uec.ac.jp, ^{††}matsubayashi.tatsushi@lab.ntt.co.jp

あらまし 位置情報を用いたソーシャルネットワーク (Location Based Social Network : LBSN) の研究は近年盛んに行われ、特にユーザの行動予測に対し、友人の行動履歴を用いることの有効性は多くの研究成果により示されて来た。従来技術では、友人の居住地関係など、空間的な距離を考慮し、予測精度の向上が示されてきた。本研究ではさらに、友人の行動類似性を考慮した行動予測手法の検証を行う。行動予測手法の検証には、2つの SNS チェックインデータを用い、個人の行動履歴の類似性に加え、友人間の共起性に着目し、訪問地予測精度の検証を行った。

キーワード 位置情報、確率モデル

1. はじめに

近年スマートフォンの普及や、Twitter や Facebook や Foursquare などの SNS サービスや、Pokemon GO などの GPS を利用したゲームの登場により、位置情報のデータが蓄積されるようになってきた。それ故、情報配信や情報ナビゲーションなどのサービスでは、ユーザの位置情報を利用した技術が非常に注目を集めている。例えば、訪日外国人などの観光客向けに、旅行者の行動ログや、趣味嗜好、所有してある観光データを分析して、旅行者の位置や属性に合わせた観光コンテンツやおすすめスポットの情報を配信するといったユーザの行動を予測し、よりユーザのニーズに合うような情報を提供するサービス [1] がある。しかしながら、行動先を決める要因としては個人の行動履歴だけではなく、友人のロコミ情報や SNS で投稿された写真から得る情報なども挙げられ、これらと関係するユーザの友人の行動履歴もユーザの行動に影響を与えうるデータと考えることができる。

Backstorm ら [2] は、Facebook の居住地が明確な約 300 万ユーザに対し、訪問地予測の検証を行い、ロケーション間の距離と同時に、友人との居住地の距離分布を考慮することで精度が向上することを示している。特に、25 マイル以内のロケーション予測では、多くの友人がいるユーザに関する予測性能の向上が見られる。Cho ら [3] では、Gowalla と Brightkite の SNS データおよび携帯電話のユーザデータを用いて、ユーザの周期的な行動分析と、友人関係に着目した訪問地予測の検証を行っている。ベースライン手法である Most Frequented Location Model (MF) に対して、曜日時間帯の周期性を考慮した Periodic Mobility Model (PMM) では 5% 程度予測精度が上がることを示され、更に、友人の影響を考慮した Periodic and Social Mobility Model (PSMM) では対数尤度と推定場所の空間的な距離の誤差で精度が上がることを示している。Cho らは、「ユーザの訪問地予測に対して、予測性能の 50 ~ 70% はユーザ自身の時間帯や曜日による周期性として解釈が可能であり、さらに予測性能の 10 ~ 30% は、友人からの影響力だと結

論付けることが可能だ」と述べている。また Scellato ら [4] でも、家からの距離が近いような場所は友人の影響を受けにくい、距離が遠い場所は友人の影響を受けやすいとしている。

前述のように、従来研究の結果より、ユーザの訪問地予測には友人の行動情報を用いることは非常に有益であり、さらに空間的に近い行動を行っているユーザの行動情報を用いるのは有効であることが示されている。しかしながら、従来の研究では、個人の行動予測モデルに友人の影響を反映させるといった事を行う際に、居住地や予測地の空間的な距離は考慮しているものの、友人との行動の類似度や面会頻度や付き合いの長さから想定されるような親密度合いは十分に考慮されていない。そこで、我々はユーザの行動予測において、個人の行動履歴と行動範囲の類似性に加え、友人間の親密度を考慮した訪問地予測精度の検証を行う。具体的には、ユーザ自身の行動の一次マルコフ遷移行列と、友人の行動の一次マルコフ遷移行列を、親密度重みを考慮した線形和モデルを提案し、予測性能の検証を行う。

2. 友人の影響力

本章ではまず、ユーザ自身の行動履歴による予測確率と、ユーザの友人の行動履歴による影響を考慮した予測確率の違いによる、友人の影響力に関する検証を行った。

2.1 使用データ

具体的には [3], [4] でも使われている、Stanford 大学が公開しているチェックインデータ^(注1)より、位置情報が付与された2つの SNS サービス、Gowalla と Brightkite の位置情報付きのチェックインデータを使用する。データの詳細は図 1 に示す。本研究ではまず日本国内のログのみを抽出し、100 以上のログを有するユーザのみ (Gowalla, Brightkite それぞれ 109 人と 544 人) を抽出した。なお、同一場所に連続でとどまり続けている履歴は同一場所遷移ログとしてデータに含めてある。図 2 はログ数ごとのユーザの分布図である。この図より、どちらの

(注1): Stanford Large Network Dataset Collection: SNAP <https://snap.stanford.edu/data/>

データもログ数が少ないユーザーが多いことがわかる。横軸は1ユーザあたりのログ数で、縦軸は横軸のログ数を保有しているユーザの数を表している。また、学習データとテストデータの比率は80:20で時系列順に抽出した。つまりテストデータの方が最新の日付のデータとなる。なお、正答率はそれぞれ移動先の確率の上位1つの正答率 (p@1[%]) と上位5つの正答率 (p@5[%]) を調べて評価する。

User ID	年月日	時間帯	曜日	位置ID	緯度	経度	住所
9853	2010/03/09	20:13:23	Mon	53456	38.----	139.----	東京都XX
45322	2010/04/09	21:45:56	Thu	44546	40.----	139.---	東京都XX
45323	2010/05/09	23:45:24	Wed	12432	40.----	139.---	東京都XX

User ID	友人ID
9853	45322
9853	45322
45322	45323

図1 使用したデータの例

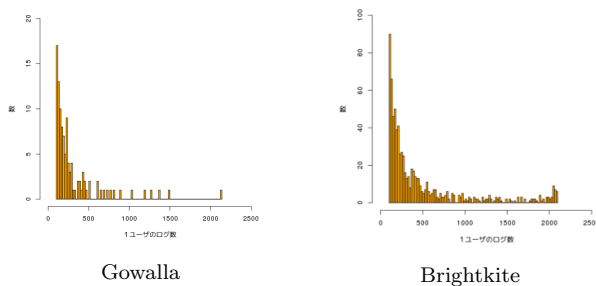


図2 Gowalla と Brightkite のユーザーのヒストグラム

2.2 マルコフモデルによる遷移確率と友人の影響力

まず、友人の影響力の定量的な検証のために、友人の行動履歴の影響を考慮した予測確率の検証を行った。

Choら[3]で言及されているように、個人の行動を予測するためには個人の周期的な行動が重要である。個人の周期性を求め手法としては一次マルコフ遷移を用いたものが幾つかあり[5],[6]、曜日、時間帯を考慮した複数の他項分布の重み付き線形和を用いた手法[6]では25~45%と、非常に高い予測精度を達成している。本稿でも同様に1次マルコフの遷移モデルを拡張し、自身の行動履歴とユーザの友人の行動履歴を線形和で表現し、重みを変えることによって検証を行った。具体的には、場所aから場所bへのユーザiの遷移確率を M_{ab}^i とし、遷移行列 X_{ab}^i を下記のように与える。

$$X_{ab}^i = (1 - \omega)M_{ab}^i + \frac{\omega}{|\mathcal{F}_i|} \sum_{j \in \mathcal{F}_i} M_{ab}^j \quad (1)$$

ここで \mathcal{F}_i はユーザiの友人の集合で、 $|\mathcal{F}_i|$ はユーザiの友人数である。また、 ω は友人の履歴の影響力の重みで $0 \leq \omega \leq 1$ である。この時、 X_{ab}^i および M_{ab}^i は下記のように正規化される、

$$\sum_b X_{ab}^i = 1, \text{ and } \sum_b M_{ab}^i = 1 \quad (2)$$

式(1)でユーザの行動を予測した結果が表1,2と図3である。

図3の横軸は ω の値を表し、縦軸は ω の値毎のp@1[%],p@5[%]の正答率を表している。図3と表1,2より、Gowallaのp@1[%]以外は正答率が増加し、どちらのデータも $\omega = 0.1$ から ω の値が上がるにつれて正答率が下がっていることがわかる。Gowallaは友人の影響を加えたことによって、p@1が0.06パーセント減少し、 ω の増加に伴い、 $\omega = 0.9$ では2パーセント減少している。p@5[%]では全ての ω の値において、正答率が0.3パーセントから1.3パーセントほど上昇している。Brightkiteではp@1[%]が $\omega = 0.1$ から0.14パーセントほど正答率が上がり、 $\omega = 0.9$ では1.1パーセントほど減少している。p@5[%]もp@1[%]と同じような正答率の変化をしている。また、どちらのデータ共に $\omega = 1$ の時に急激に正答率が下がっているのは友人の行動履歴しか学習していないためである。

表1 式(1)で友人の影響を考慮したGowallaの行動予測精度の結果

友人の影響	p@1[%]	p@5[%]
遷移行列のみ	12.45	22.08
$\omega = 0.1$	12.39	23.37
$\omega = 0.2$	12.33	23.34
$\omega = 0.3$	12.11	23.34
$\omega = 0.4$	11.80	23.34
$\omega = 0.5$	11.60	23.31
$\omega = 0.6$	11.52	23.30
$\omega = 0.7$	11.31	23.20
$\omega = 0.8$	10.61	23.04
$\omega = 0.9$	10.16	22.34
$\omega = 1.0$	1.84	3.24

表2 式(1)で友人の影響を考慮したBrightkiteの行動予測精度の結果

友人の影響	p@1[%]	p@5[%]
遷移行列のみ	17.15	20.46
$\omega = 0.1$	17.29	20.77
$\omega = 0.2$	17.28	20.78
$\omega = 0.3$	17.27	20.78
$\omega = 0.4$	17.00	20.77
$\omega = 0.5$	16.98	20.77
$\omega = 0.6$	16.90	20.75
$\omega = 0.7$	16.78	20.73
$\omega = 0.8$	16.63	20.66
$\omega = 0.9$	16.06	20.44
$\omega = 1.0$	0.27	0.44

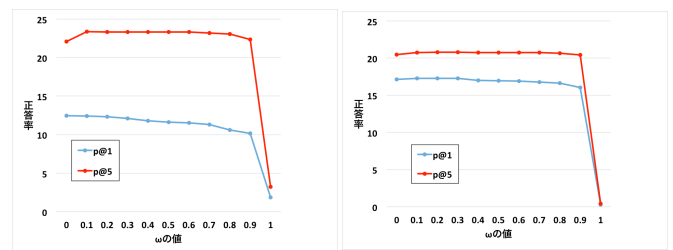


図3 式(1)での ω の値ごとの正答率

図4, 図5は両データの $\omega = 0.1$ と $\omega = 0.9$ の時のユーザごとの正答率の分布である。これらの図の各点はユーザを表し、横軸はそれぞれ $p@1[\%]$ と $p@5[\%]$, 縦軸の $p@1(\omega = 0.1)[\%]$, $p@5(\omega = 0.1)[\%]$ というのは、それぞれの ω の値を用いて式(1)で求めた $p@1[\%], p@5[\%]$ を表している。図4と5で対角線上にある点は、友人の影響を与えても正答率が変化していないユーザを表しているので、対角線上の上下に分布しているユーザは正答率が変化していることを表している。全ての図において、対角線の上側にあるユーザは $p@1[\%], p@5[\%]$ の値が共に10パーセントから20パーセントである事が多く、正答率の上昇幅も10パーセント以内で収まっているの者が多い。 $\omega = 0.1$ と 0.9 の違いを見比べると、 $\omega = 0.1$ から 0.9 になった時に、対角線の下に位置するユーザが増えている事がわかる。この事は自己遷移列のみでは正答率の高くないユーザが友人の影響を与えることで正答率が下がっている事が多いということを示している、友人の影響を受けているほとんどのユーザの正答率が下がっている事がわかる。友人の影響を加えることで、正答率が大きく下がったり、0に近い値になっているユーザがいるが、理由として、学習データのうち、連続して同じ場所へ訪問しているデータの割合が高い事や、そのような友人のデータが多い事が挙げられる。自身の学習データやテストデータに同一場所訪問回数が多いと、自己遷移のみを学習すると正答率が高いが、友人のログの影響で、そのような場所が当たらなくなってしまう、正答率が大きく下がってしまう。友人に同一場所訪問回数の多いユーザが多いと、その友人が習慣的に訪問している場所(例えば最寄りの駅や職場など)を、選択する確率が高くなってしまふ。次に、ユーザーごとの傾向を見るために、友人の影響を加える事によって $p@1$ の値が増加、減少したユーザの詳細を調べた。両データの友人の影響を加えた $p@1$ において一番値が高くなった $\omega = 0.1$ の時についてその様なユーザー詳細を調べたものが表3と表4である。

表3 友人の影響を考慮することで $p@1$ の値が変化した Gowalla のユーザ情報

ユーザのログ数	$p@1[\%]$	$fp@1[\%]$ 友人影響	友人数	友人のログ数	$p@1(\omega = 0.1) - p@1$
229	4.36	13.11	3	309	8.73
281	3.56	10.68	12	858	7.12
214	2.34	7.01	11	1610	4.67
248	4.03	8.06	2	1049	4.03
117	8.55	4.27	4	115	-4.27
144	6.94	0	27	1556	-6.94
236	16.94	8.45	4	45	-8.47
323	44.89	7.74	42	7089	-37.15

表4 友人の影響を考慮することで $p@1$ の値が変化した Brightkite のユーザ情報

ユーザのログ数	$p@1[\%]$	$fp@1[\%]$ 友人影響	友人数	友人のログ数	$p@1(\omega = 0.1) - p@1$
176	0	5.71	18	11116	5.71
280	7.14	12.50	32	13927	5.36
307	1.64	6.56	51	21118	4.91
245	4.08	8.16	6	2427	4.08
256	39.95	38.21	121	48721	-1.74
1677	18.86	16.98	13	1746	-1.89
1249	12.75	9.80	26	5624	-2.94
1595	8.00	2.00	10	2440	-6.00

これらの表は友人の影響を与えることで $p@1$ の正答率が増加、減少したユーザーを値が大きい順にそれぞれ4人ずつ載せたものである。背景が濃い行は正答率が増えたもので、白いものは減ったものである。Brightkite よりも Gowalla の方が $p@1[\%]$ の変化幅が大きいことがわかる。続いて、ユーザのログ

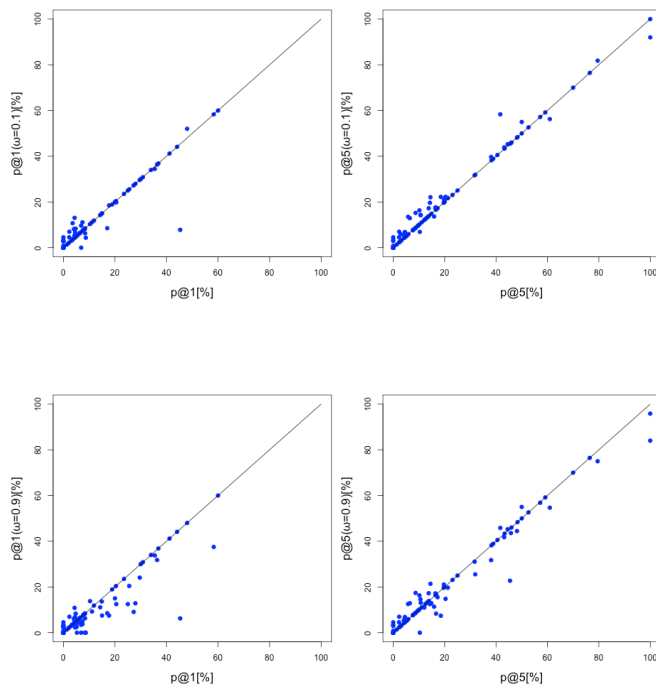


図4 式(1)で $\omega = 0.1, 0.9$ の Gowalla の正答率の分布

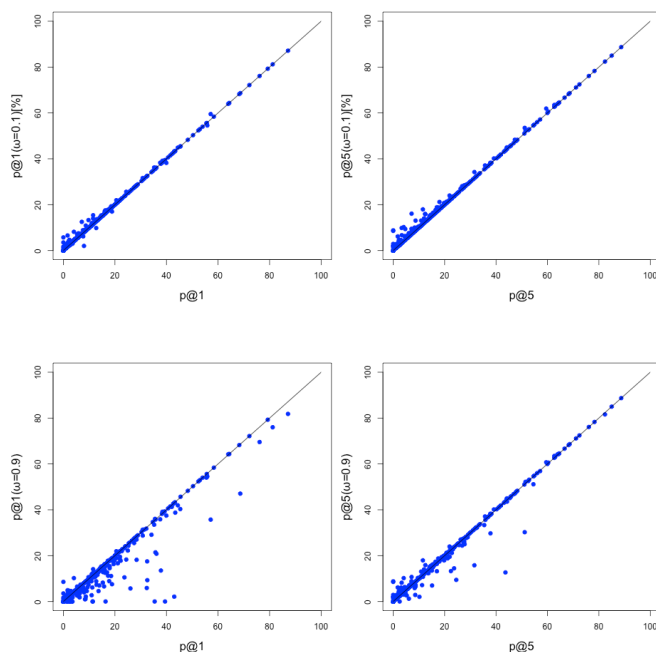


図5 式(1)で $\omega = 0.1, 0.9$ の brightkite の正答率の分布

数とその友人のログ数の割合と、正答率の変化に相関がないかを確かめるために、図6にユーザの分布をプロットした。この図の横軸はそれぞれのデータの $p@1$ と $p@1(\omega = 0.1)$ の差を

表し、縦軸はユーザごとの友人のログ数の総和 (n_flog) をそのユーザ自身のログ数 (n_log) で割ったもので、自分のログ数に対して友人のログ数の総和の方が多いほど大きな値となる。この図より目立った相関関係は見る事ができないが、強いて言えば、どちらのデータも正規分布のように見える。また、図 6 の Gowalla の左端にプロットされているユーザは、図 4 で確認される、 $p@1(\omega = 0.1, 0.9)$ で正答率が大きく下がっているユーザである。さらに Brightkite の図では横軸が 5 以上で縦軸が 50 から 60 に収まっている集団も確認することができる。次に友人の影響を考慮することで当てることができた場所に着目した。この時予測を当てる事のできた場所はゲームセンターや居酒屋などと言った友達と定期的に通うような所であり、同時期、時間帯に実際に友人と訪れている事が確認できた。また、友人の影響を考慮したことで正答率の上がったユーザと下がったユーザの位置情報の分布に着目した所、正答率の上がったユーザは友人との行動範囲が重複していて、そうでないユーザは重複していないという仮説が得られた。その例を顕著に表しているものが図 7 と図 8 である。赤いプロットが対象にしたユーザのチェックインログの位置を示し、青いプロットがそのユーザの友達 (複数人) のチェックインログの位置を示す。横軸は経度で縦軸は緯度だが、その値はユーザのプライバシー保護のため表示していない。正答率が下がっているユーザは自分のチェックインした位置以外に、多くの友人のチェックインログの位置があり、自分が行ったことの無い位置の影響を受けていることがわかる。逆に、正答率の上がったユーザはその友達と行動範囲が類似している事がわかる。

3. 実験と考察

3.1 共起度を用いた友人影響度

前節で議論したように、行動範囲が類似した友人の行動の影響力を考慮する事が重要になりうる。その事を検証するために、まずは行動の類似性の指標として、下記の幾つかの共起度指標を考える。

$$\text{Jaccard 係数} = \frac{|X_i \cap X_j|}{|X_i \cup X_j|} \quad (3)$$

$$\text{Simpson 係数} = \frac{|X_i \cap X_j|}{\min(|X_i|, |X_j|)} \quad (4)$$

$$\text{ダイス係数} = \frac{2|X_i \cap X_j|}{|X_i| + |X_j|} \quad (5)$$

$$\text{コサイン} = \frac{|X_i \cap X_j|}{|X_i| |X_j|} \quad (6)$$

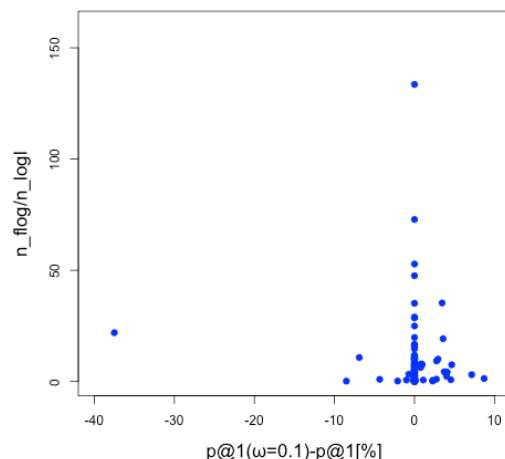
ここで X はユーザ i のチェックイン箇所の集合で、 $|X_i|$ はチェックインの共起数 (場所数) とする。

ユーザ i とユーザ j の共起度を s_{ij} とした時、ユーザ i の遷移確率を X_{ab}^i は下記の用に定義される、

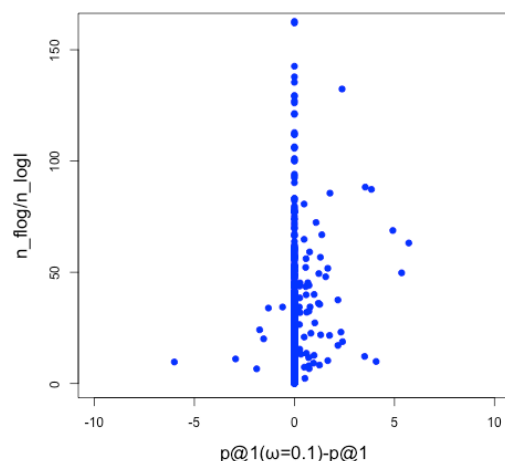
$$X_{ab}^i = (1 - \omega)M_{ab}^i + \frac{\omega}{|C_i|} \sum_{j \in \mathcal{F}_i} s_{ij}M_{ab}^j \quad (7)$$

この時、正規化は式 2 を満たし、さらに下記を満たすようにとる。ここで C_i は正規化定数である、

$$\sum_{j \in \mathcal{F}_i} s_{ij} = C_i \quad (8)$$



Gowalla



Brightkite

図 6 $fp@1$ と $p@1$ の差と友人ログ数と自己ログ数の割合の分布図

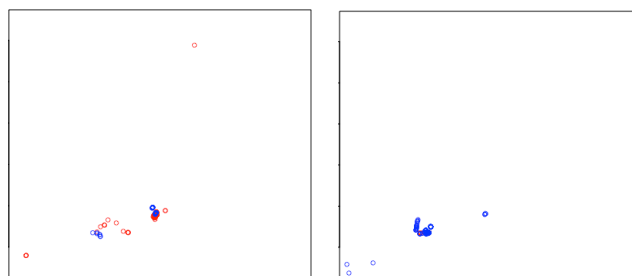


図 7 $p@1$ で上がったユーザとその友人の行動範囲

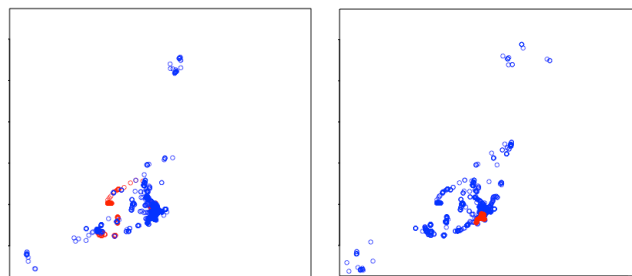


図 8 $p@1$ で正答率が下がったユーザとその友人の行動範囲

それぞれの共起度は 0 から 1 の値をとり、1 に近いほどユーザ i とユーザ j の行動範囲が類似していることを表す。Jaccard 係数はユーザーの訪問したロケーションのユニーク数に差があると値が小さくなってしまいうという短所がある。この場合ユーザと友人の訪問履歴が一致する場合が多くても、ユーザに対して友人の訪問場所のユニーク数が多いと、共起度の値が小さくなってしまふ。Simpson 係数では分母に最小値をとるようにしているので、2 者間の訪問地のユニーク数に差があっても、このように値が小さくなるという事は無くなる。しかし、ユーザが友人の訪問場所のユニーク数のどちらか一方が小さいと高い値が出やすくなるという短所もある。この章ではユーザと友人の行動範囲の類似度を考慮した式 (7) での $p@1, p@5$ の検証を行う。

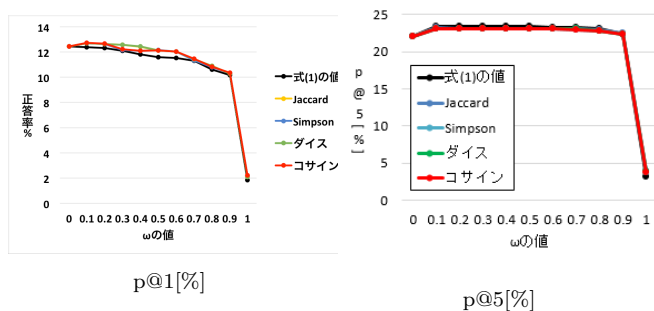


図 9 式 (7) の Gowalla の行動予測精度の結果

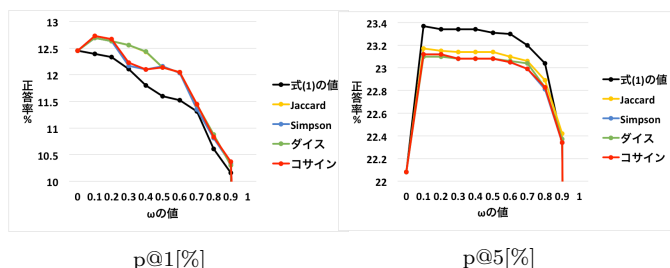


図 10 図 9 のそれぞれの係数の変化に着目したもの

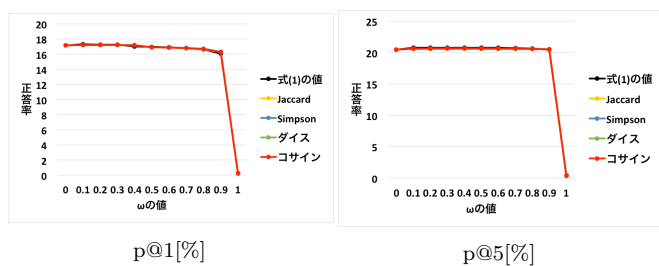


図 11 式 (7) の Brightkite の行動予測精度の結果

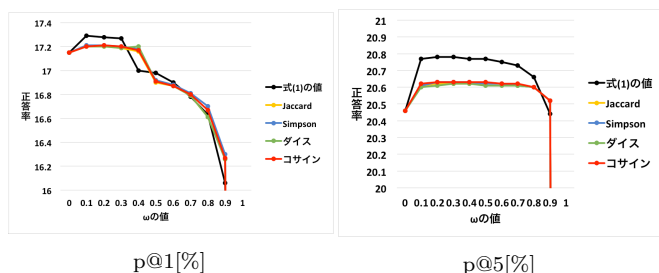


図 12 図 11 のそれぞれの係数の変化に着目したもの

3.2 ユーザーと友人の友人の行動範囲を考慮した検証

まず、式 (7) の検証を行った。検証に用いたデータ、条件は 3 節で用いたものと同様で、それぞれのデータで $p@1$ と $p@5$ を求めた。共起度の指標には式 (3) ~ (6) を用いてそれぞれ値を比較した。なお、これらの指標の計算には、同一訪問場所の回数は、全て 1 回として計算しており、訪問頻度は反映させていない。図 9, 図 11 に両データの結果を示す。式 (1) と比べて、Gowalla の $p@1$ [%] のみ正答率 0.36 パーセント増加し、他は全て 0.1 から 0.2 パーセントほど減少した。式 (1) の友人の影響で唯一正答率が下がったものだけが、今回は上昇しているが、これは Gowalla の $p@1$ [%] のみが行動範囲が類似した友人の行動の影響力を考慮することが重要という仮説があてはまるケースであると考えられる。図 10, 12 ではそれぞれの係数ごとの正答率の変化を調べるために、図 9, 11 の縦軸の範囲を狭めたものである。図での各係数の値は線が重なっていることが多く、係数ごとの違いは大きく観測されない。図 13 と 14 は $\omega = 0.1, 0.9$ のユーザごとの正答率の分布を表している。横軸は自己遷移行列によるものみの正答率で、縦軸はそれぞれの係数による共起度を与えたときの正答率である。全体的に見て、図 4 と 5 と比べて、対角線の上側と下側にプロットされているユーザが対角線によっている事がわかる。これは共起度付きの友人の影響を与えることで式 (1) で正答率の下大きく下がったユーザの正答率は増加しているが、式 (1) で正答率が大きく増加したユーザの正答率を下げてしまっていることを意味する。特に図 4 と 5 で確認された、対角線上から大きく離れたところにプロットされているユーザが存在したが、図 13 と 14 では対角線上に近い所に位置している。この事から、自身や友人の学習データやテストデータに同一場所訪問ログを多く持つユーザでは、共起度で友人の影響の負の効果を打ち消すことができていると考えることができる。また、ほとんどのプロットが重なっていて、同じ色になっているので係数での違いはあまり観測されないが、Gowalla の $\omega = 0.1$ の $p@5$ [%] ではDice係数が Brightkite の $\omega = 0.1$ の $p@1$ [%] と $p@5$ [%] では Simpson 係数が、他の係数よりも値が良くなっているユーザがわずかにいることが観測される。

3.3 共起度を加えることによって正答率が上昇したユーザー

表 5 は式 (1) と式 (7) での $fp@1$ と $p@1$ の値が変化したユーザの数を表したものである。正、負というのは $fp@1-p@1$ の符

表 5 式 (1) と式 (7) で $fp@1-p@1$ の値が変化したユーザの数

データ	式 (1) 正 [人]	式 (7) 正 [人]	式 (1) 負 [人]	式 (7) 負 [人]
Gowalla	18	16	7	0
Brightkite	57	49	7	12

号を表し、正だとそのユーザにとって、遷移行列のみの予測よりも精度が上がっていて、負だとその逆になる。式 (1) と式 (7) の違いについて、Gowalla は式 (7) の方が式 (1) より精度が上がっているユーザが 2 人ほど少ないが、下がっているユーザは 0 人になっていて、正答率が変動したユーザの数は式 (1) の方が 9 人多く、図 9 から、全体の正答率はわずかに高い。この事

から Gowalla では共起度を影響させることで、式 (1) の友人の影響の負の効果を打ち消して、自己遷移行列のみの正答率と比べて、全体的な精度を上げていることがわかる。Brightkite は式 (7) の方が正答率の予測精度が上がっているユーザが 8 人少なく、予測精度が下がっているユーザは 5 人多くて、正答率が変動したユーザは 6 人減っている。図 11 より、全体の正答率はわずかに下がっている事がわかるので、Brightkite では共起度を影響させることで式 (1) の正の影響を打ち消してしまったと考えられる。表 6 と表 7 は $\omega = 0.1$ での式 (1) の正答率と、式 (7) の正答率を比べて増加率と減少率が大きいユーザ上位 4 人を並べたものである。表の $fp@1[\%]$ は 4 つの係数の確率の中で最も値の大きいものを選んでいる。表の背景を色塗ってあるユーザは表 3,4 にも乗っていたユーザと一致しているものである。全体的に見て表 3,4 と比べて上位 4 ユーザの正答率は小さく、下位 4 ユーザの正答率の値は大きくなっている。このことは図 13,14 の議論でのユーザのプロットが対角線上に近くなった事と関連性があると考えられる。Gowalla の上位 4 人のうち 3 人のユーザが表 3 のユーザと一致しており、これらのユーザは正答率は式 (7) のほうが下がっているので、自己遷移による正答率の効果を弱めてしまったことがわかる。Brightkite ではこの表の中では 1 行目のユーザのみ式 (7) で正答率が上昇しているが、このユーザの訪問場所は友人の訪問場所と一致している事が多かった。また、式 (7) で正答率の減少率が大きいユーザは式 (1) で減少率が大きかったものとあまり一致しなかったが、これらのユーザーに共通する特徴は見当たらなかった。

表 6 共起度を考慮することで $p@1$ の値が変化した Gowalla のユーザ情報

ユーザのログ数	$p@1[\%]$	$fp@1[\%]$	共起度影響	友人数	友人のログ数	$fp@1-p@1$
229	4.36	10.87		3	309	6.51
113	0	4.55		1	89	4.55
248	4.03	8.33		2	1049	4.33
423	10.71	14.29		23	8142	3.57
323	44.89	44.89		42	7089	0
236	16.94	16.94		4	45	0
144	6.94	6.94		27	1556	0
117	8.55	8.55		4	115	0

表 7 共起度を考慮することで $p@1$ の値が変化した Brightkite のユーザ情報

ユーザのログ数	$p@1[\%]$	$fp@1[\%]$	共起度影響	友人数	友人のログ数	$fp@1-p@1$
245	4.08	10.87		6	2427	6.78
176	0	5.71		18	11116	5.71
280	7.14	10.71		32	219327	3.57
307	1.63	3.27		51	21118	1.64
974	8.71	7.18		52	19563	-1.53
320	3.13	1.56		42	15356	-1.56
227	22.22	20.00		25	15084	-2.22
536	8.41	0.93		47	19470	-7.48

4. まとめと展望

位置情報ログを利用したユーザの行動予測に友人の影響を与えたモデルの検証を行った。友人の影響を考慮することによる位置予測で、正答率が変化するユーザーの特性として、行動範

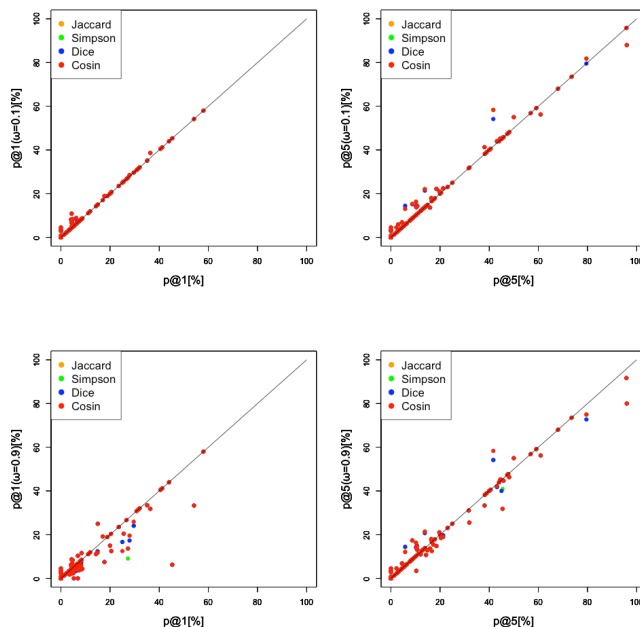


図 13 式 (7) で $\omega = 0.1, 0.9$ の Gowalla の正答率の分布

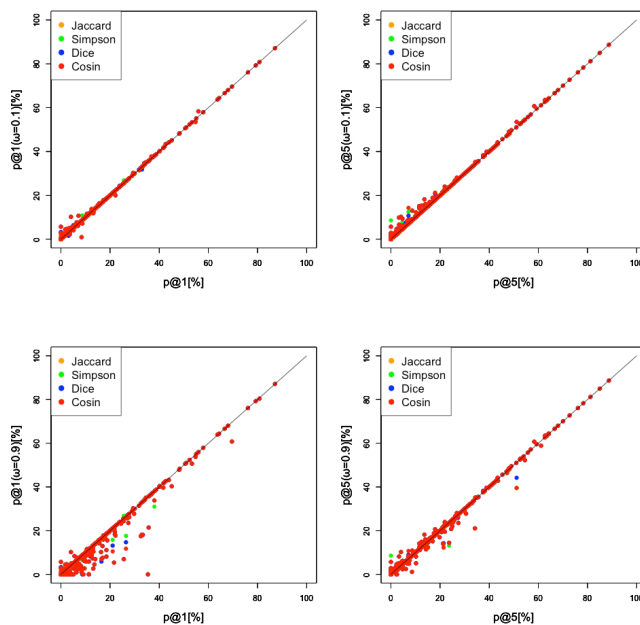


図 14 式 (7) で $\omega = 0.1, 0.9$ の brightkite の正答率の分布

圏の類似度に着目した。友人の影響を与えることで最大1パーセント正答率が増加した。自身の行動履歴と友人の行動履歴を線形和で表現したものと、おおむね全体的な正答率の上がり、自身と友人との行動履歴の行動範囲の共起度を影響させ重み付けさせることで、特定のデータでは正答率を上げることができた。しかし、今回検証した、友人のログや共起度を考慮する手法だと、Gowalla,Brightkite のテストデータの対象になったユーザ数のうち22,12パーセントほどしか、式の影響を受けていなかった。より現実に即し、行動予測に有益な友人の性質をモデル化に影響させる必要がある。今後の方針として、今回は考慮していなかった、友人の質をグラフのクラスタ係数やマルチエージェントでシミュレーションする事などで定義し、モデルに反映させるような事や、深層学習を用いて友人情報を絡めた、各ユーザーの予測位置の重み付けをするというような事を検討している。また、訪問先の意思決定に、友人の影響を受け易いユーザーや場所の特徴を抽出できると、友人情報を利用した情報推薦アプリなどに応用できるので、そのような特徴を抽出するような事も検討している。

文 献

- [1] 熊谷 雄介, 今井 良太, 松林 達史, 佐藤 吉秀, 堀岡 力. .”非負値複合テンソル因子分解を用いた訪日外国人観光客の回遊行動分析.”, 信学技報 115.112 ,2015
- [2] Backstrom, Lars, Eric Sun, and Cameron Marlow. “Find me if you can: improving geographical prediction with social and spatial proximity.” Proceedings of the 19th international conference on World wide web (WWW). ACM, 2010.
- [3] E. Cho,S. A. Myers,and J. Leskovec. “Friendship and mobility: user movement in location-based social networks.” Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1082–1090.ACM, 2011.
- [4] S. Scellato, A.Noulas, R.Lambiotte, and C.Mascolo, “Socio-spatial properties of online location-based social networks.”In ICWSM’11, 2011.
- [5] D. Ashbrook, and S. Thad, “Using GPS to learn significant locations and predict movement across multiple users.” Personal and Ubiquitous Computing, 7(5),pp.275–286, 2003.
- [6] 林 亜紀, 松林 達史, 澤田 宏 “位置情報を利用した情報配信のための習慣度算出手法” 日本データベース学会和文論文誌 13-J(1),pp. 64–71.DBJSJ, 2014.